

<b>Devi Ahilya University, Indore, India Institute of Engineering &amp; Technology</b>			<b>III Year B.E. (Computer Engineering)</b>				
<b>Subject Code &amp; Name</b>	<b>Instructions per Week</b>		<b>Hours</b>	<b>Credits</b>			
<b>5CERG3 Data Mining</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>Total</b>
<b>Duration of Theory Paper: 3 Hours</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>4</b>

### Learning Objectives:

1. Ability to understand the role of data mining in knowledge discovery process.
2. To familiarize with various data mining functionalities and how it can be applied to various real-world problems.
3. To learn about finding data characteristics and evaluating the outcome of data mining process.
4. To familiarize with various machine learning algorithms used in data mining.

### Prerequisites:

The students are required to have some basic knowledge about Data structures and Databases.

## COURSE CONTENTS

### UNIT-I: Introduction

Data Mining: Overview, Type of data mined, Functionalities, Technology used, Target applications and challenges; Data Features: Attribute types, basic statistical description, measuring data similarity and dissimilarity.

### UNIT-II: Data Pre-processing and Data Warehouse

Data Pre-processing: Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Discretization; Data Warehouse: Basic concepts, Design and Usage, Implementation.

### UNIT-III: Frequent Pattern and Association rule Mining

Basic concepts, Pattern Mining: Apriori algorithm, FP-growth Algorithm; Generating association rules, Pattern evaluation methods, Multi-level and multi-dimensional pattern mining.

### UNIT-IV: Classification

Classification: Basic Concepts, Decision Tree Classifier: Decision tree algorithm, attribute selection measures, Bayes Theorem and Naive Bayes Classifier, Model Evaluation: Holdout Method, Cross-Validation, evaluation metrics, confusion matrix.

### UNIT-V: Cluster analysis and Advance Topics

Clustering: Introduction, Types of clustering; Partition-based clustering: K-Means, K-Medoids; Density based clustering: DBSCAN, Clustering evaluation.

## Course Outcomes:

Students who have completed this course should be able to:

1. Apply data mining functionalities on real world problems and datasets.
2. Have some knowledge about the couple of data mining tools and how they can be used for large data.
3. They would be able to find the characteristics of given data and may identify presence of outliers, if any.
4. The course would help them to pursue some advance course on data science and may help in subjects like Big Data, AI etc.

**Assessment methods of all of the above: quizzes, exams, assignments.**

Co No.	CO	PO
CO1	Learn about different types of Data ,Features and Application	PO1,PO9,PO11
CO2	Analyse various Data Pre-processing approach	PO2,PO3,PO5 ,PO11,
CO3	Learn about Frequent pattern and Association rules .	PO2,PO4, PO6
CO4	Learn about Classification method and Evolution measures..	PO2,PO4,PO5,
CO5	Learn about Clustering method and Evolution measures.	PO4,PO5 , PO12

## CO PO Relationship

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3								2		3	
CO2		2	3		3						3	
CO3		3		3		2						
CO4		3		3	3							
CO5				3	3							3

## Books Recommended:

[1] Han, Kamber and Pi, Data Mining Concepts & Techniques, Morgan Kaufmann, 3<sup>rd</sup> Edition, India, 2012.

[2] Mohammed Zaki and Wagner Meira Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014.

[3] Z. Markov, Daniel T. Larose Data Mining the Web, Jhon wiley & son, USA, 2007.

[4] Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, 2nd Edition, 2011.

[5] Sam Anahory and Dennis Murray, Data Warehousing in the Real World, Pearson Education Asia, 2000.

[6] W. H. Inmon, Building the Data Warehouse, 4th Ed Wiley India, 2005.